Review of LQCD-II Computing at BNL

2016 DOE Annual Progress Review
JLAB
June 28-29, 2016

Robert Mawhinney
Columbia University
Co-Site Architect for BNL

BNL Computers used for QCD

12k node QCDSP, 600 GFlops, 1998-2005
RIKEN-BNL Research Center (RBRC)

2 ×12k node QCDOC, 20 TFlops, 2005-2011
DOE for USQCD and RBRC

2k node RBRC BGQ, 400 TFlops, 2012-
1k node BNL BGQ, 200 TFlops, 2012-

3k nodes RBRC/BNL BGQ, 600 TFlops, 2012-
0.5 k nodes USQCD BGQ, 100 TFlops, 2013-

BGQ Computers at BNL

USQCD half-rack (512 nodes)

2 racks of DD1 RBRC

1 rack of DD2 BNL

BNL Participation in USQCD Post-BG/Q

- The BNL BGQ computers continue to serve the needs of USQCD, BNL and the RBRC
  * The USQCD BG/Q has run jobs over 96% of the wall clock time since July 1, 2015
  * BNL's Information Technology Division (ITD) has managed this resource, along with QCDSP, QCDOC, BG/L and BG/P from 1998-2016.
- BNL has a large committment to QCD and LQCD
  * The RHIC program, ATLAS, ...
  * The largest Tier 1 ATLAS computer facility
  * Many LQCD theorists in the Physics Department
  * LQCD theorists (and others) at the RBRC.
- BNL is very interested in continuing to support USQCD, by hosting the best new architectures for LQCD calculations
  * There is first-rate expertise at BNL, through the operation of the RHIC-ATLAS Computing Facility (RACF).
  * There is also a new initiative at BNL to coordinate computing lab-wide, the Computational Science Initiative.

Computational Science Initiative (CSI)

- Established in December 2015 to provide an integrating umbrella for all computer science and applied mathematics research and services.

- Provide a focal point for leading data science research
- Educate the next generation of expert data scientists
- Translate data science research advances into tools and expertise that lead to measurable scientific progress
- Provide a State of the Art Data, Computing and Network Infrastructure to meet research and operational requirements for BNL and its Partners

COMPUTATIONAL SCIENCE INITIATIVE

Kerstin Kleese van Dam, Director*

Michael Ernst, Deputy Director
Robert Harrison, Chief Scientist
Lauri Peragine, Administrative Assistant

BNL Scientific Data & Computing Center (BSDCC)

Eric Lancon, Director

Incl. *RACF***, Institutional Computer, C3D Systems,…

Computational Science Laboratory (CSL)
Nicholas D' Imperio, Director

Collaborative Laboratory for Advanced Algorithm Development and Optimization - bringing together HPC, Math and Domain Science Expertise

Computer Science & Mathematics (CSM)

Barbara Chapman, Technical Director

Data Science, HPC, Applied Mathematics Research

Center for Data Driven Discovery (C3D)
Kerstin Kleese – Director (interim)

Multidisciplinary Center for the Development, Deployment and Operation of Data Driven Discovery Services

• Reports to Robert Tribble, Deputy Director
** Matrixed from the Physics Department
*** Provides enabling capabilities to CSI

Approved: _____

Kerstin Kle e

Brookhaven Science Associates

BROOKHAVEN NATIONAL LABORATORY

BNL Scientific Data and Computing Center

- World class scientific computing center operation provided by the RACF team for all scientific computing at BNL and for its partners.
- High Throughput Computing
- High Performance Computing
- Data Intensive Computing
- Fast Data Storage
- Petascale Data Archive
- Fast Networking

Michael Ernst
(just retired)

Eric Lançon

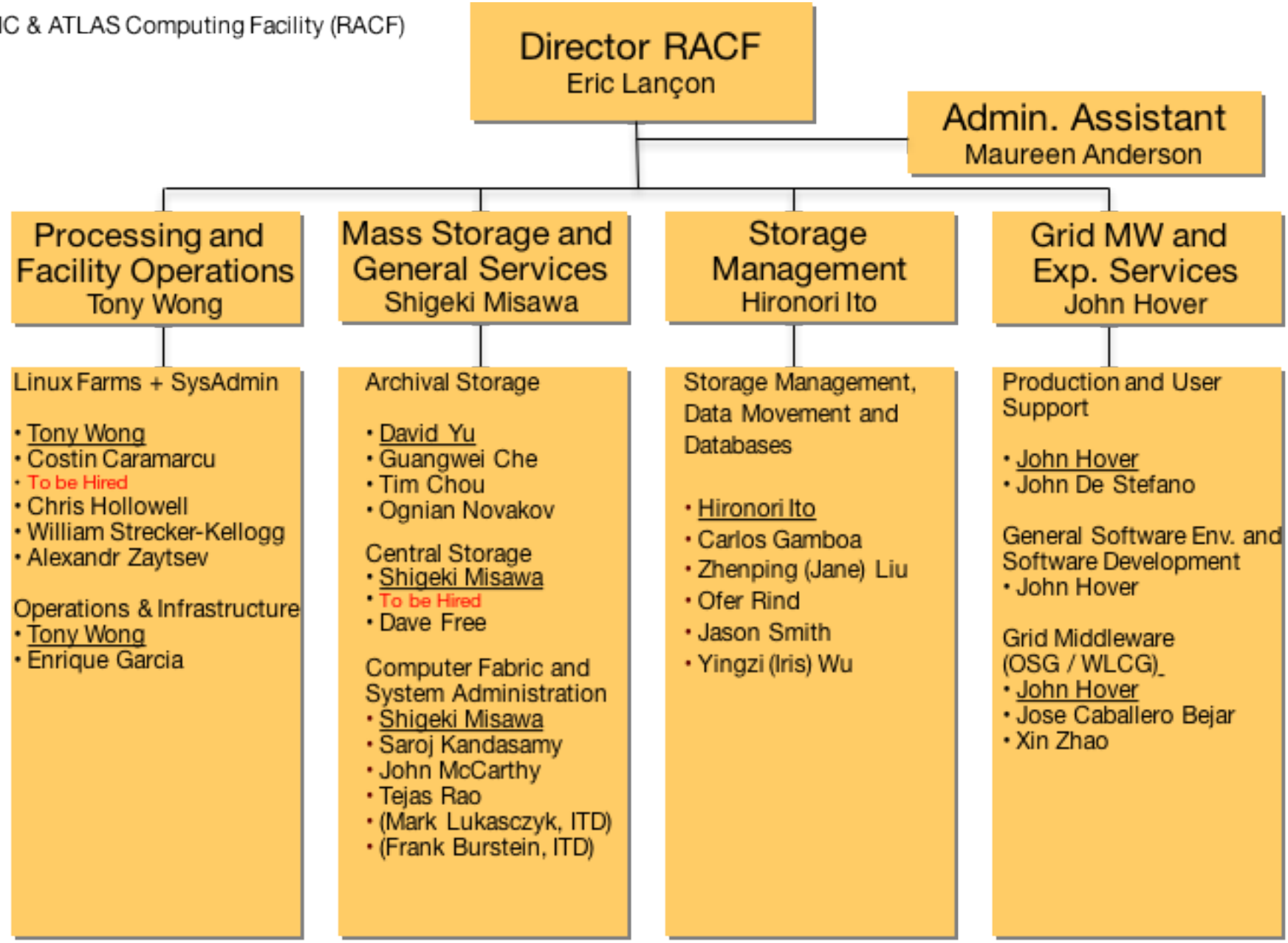
BNL Computing Support

- 2001 - Ongoing - RHIC and LHC ATLAS Computing - Largest Tier 1 Center in the world - High Throughput Data Analysis
- 2007 IBM BlueGene/L - No. 5 in the World
- 2009 IBM BlueGene/P - No. 250 in the World
- 2011 IBM BlueGene/Q - No. 6 Graph Systems in the World
- 2016 BNL Network Upgrade - connecting NSLS II, CFN and Computing Center
- 2016 Institutional Cluster 200 Node CPU/GPU + PBs of storage
- 2016 Novel Data-Intensive Architecture Testbed Facility

Overview of the RACF

- Providing centralized computing services to the RHIC experiments for 20+ years
- Expanded in the late 1990's to act as the Tier 1 computing center for ATLAS in the US
- Under CSI: goal is to unify all scientific computing operations into one organization
  * Currently 27 FTE's
  * 15,000 square foot datacenter, expanding into new space for HPC activities
  * 70 PB of data stored on ~60k tapes in 8 Oracle/StorageTek SL8500 tape libraries
  * 6 GPFS filesystems with ~6 PBytes of total storage
  * ~2,000 processor systems providing ~50k logical CPU cores
- Purchase order for multidisiplinary BNL Institutional Cluster has been released.

RHIC & ATLAS Computing Facility (RACF)



Names underlined: Section Leaders 29 Individuals

June, 2016

Computational Science Laboratory

- Collaborative Center in support of advanced Numerical Modeling
- User Support
- Code Optimization
- Scalable Mathematical Libraries
- Development of scalable Algorithms
- Benchmarking

- NVIDIA GPU Research Center 2016

CSI has, and will continue, to provide support for the LQCD user community with experts in computer science and QCD. Currently has 2 members with QCD experience, Meifeng Lin and Zhihua Dong.

Nick D'Imperio

CSI Summary

- CSI is a new organization based on key areas of Computer Science, Applied Mathematics and Computational Science expertise across BNL.
- Off to a successful start:
  - New core capability recognized at BNL in : Advanced Computer Science, Visualization & Data
  - $15M New York State funding arriving in the next couple of months
  - New Partner in US LQCD consortium - computer systems hosting and user support
  - NVIDIA GPU Research Center awarded, New member of OpenACC

BNL Hosting of USQCD Cluster Hardware

- BNL CSI strongly supportive of proposal to host USQCD clusters at BNL
  * As noted in previous talks, as part of this proposal BNL will provide USQCD access to the Institutional Cluster (IC) as of mid to late summer.
  * In the proposal, FY 2017-2018 USQCD hardware is to be procured and managed coherently with other BNL computing resources, such as RHIC and ATLAS computing, by the Scientific Data and Computing Center of CSI.
  * Through this proposal, USQCD users will have access to a high throughput (up to 24 GBytes/sec) shared file system that is part of the IC.
  * The IC gives USQCD users the opportunity to run jobs on larger partitions for shorter periods of time. Can be very useful if total partition memory is important.
- This a very exciting opportunity for USQCD and BNL.
- BNL personnel, with extensive experience in running and managing the RHIC and ATLAS Tier 1 computing facility at BNL, are now supporting USQCD as a 3/1/16
  * Tony Wong: USQCD Site Manager, head of Processing and Facility Operations of RACF
  * Shigeki Misawa: USQCD co-Site Architect, head of Mass Storage and General Services of RACF

BNL Institutional Cluster

- The Institutional Cluster is the first significant hardware component in support of HPC activities at BNL
  - 108 servers with dual-socket Broadwell CPU's
  - 2 Nvidia K80 GPU's per server
  - 128 GBytes of ECC DRAM per server
  - EDR Infiniband interconnect
  - 1 PB GPFS-based storage with 24 GBytes/second of bandwidth
  - SL 7.x (or CentOS equivalent) operating system
  - SLURM workload manager
- The upgrades to the data center power and cooling infrastructure to accommodate the IC were completed in May, 2016.
- The IC order has been placed and account creation activities have begun. Early users may be on this machine by mid-summer.
- A small test machine is available for code development and testing. This is a cluster of 5 machines, with no GPU's, on loan from other RACF activities. It will be returned when the IC is available.

CSI Support for USQCD

- As for March 1, 2016, SDCC has taken over operations for the USQCD half-rack of BG/Q and the DD2 rack.
  * Support for the legacy BG/Q and DD2 from USQCD through September 2017
- Enhanced support for USQCD under the 3-site FY-straddle (JLAB, BNL, FNAL) deployment model begin proposed:
  * The integrated compute power of 40 nodes of the IC are to be dedicated to USQCD. Larger node-count jobs can also be run consistent with the integrated maximum.
  * Access to the RACF mass storage system (intermediate disk caching and tape)
  * 100-200 TBytes on the 1 PByte high-bandwidth disk system of the IC will be available to USQCD users, with 300+ TBytes of space available on the existing, lower-bandwidth BGQ file system.
  * SDCC will support the evalutation, purchase and operation of successor(s) to the BG/Q in FY17 and FY18.

Conclusions and Outlook

- BGQ half-rack and DD2 are running very reliably and no changes are planned.
  * Joe DePace and Robert Riccobono handle hardware issues
  * System administration by Zhihua Dong
- BNL Computational Science Initiative strongly supportive of proposal for USQCD clusters at BNL
  * Access to BNL IC for USQCD
  * USQCD can tap into the extensive expertise of the staff managing the RHIC and ATLAS computing
- Exciting opportunities for USQCD and BNL, with the renewed focus of BNL on all aspects of computing.
- Many thanks to Frank Quarant and Nick D'Imperio for their leadership of and work on the USQCD project at BNL for the last many years.